OPIC
OFFICE DE LA PROPRIÉTÉ INTELLECTUELLE DU CANADA

CIPO
CANADIAN INTELLECTUAL PROPERTY OFFICE

Ottawa Hull K1A 0C9

(21) (A1) 2,125,240
(22) 1994/06/06
(43) 1995/12/07

(51) Int.Cl.[5] C12P-021/00; C12P-019/34

(19)(CA) **APPLICATION FOR CANADIAN PATENT** (12)

(54) Synthetic Gene Libraries

(72) Deng, Su-jun - Canada ;
MacKenzie, C. Roger - Canada ;
Narang, Saran A. - Canada ;
Young, N. Martin - Canada ;

(71) National Research Council of Canada - Canada ;

(57) 12 Claims

ABSTRACT

A method for the simultaneous randomization of multiple and specific regions of a protein is described. This method also controls the degree to which this randomization occurs. Targeted sequences within a protein are partially randomized by the precise randomization of synthetic deoxyribooligonucleotides. Genes encoding proteins are assembled from the randomized oligonucleotides using a ligase reaction and are expressed on bacteriophage surfaces for selection of novel activities. The method can be used for randomizing antibody complementarity determining regions or framework regions for enhancing the affinity of the mutated antibodies. In one example, mutant single-chain antibodies, specific for the Salmonella serogroup B O-antigen, with 100-fold better binding, relative to the wild-type, were isolated using the invention.

SYNTHETIC GENE LIBRARIES

Field of Invention

This invention relates to a method for the controlled randomization of a gene encoding a protein at a selected number of sites along the gene and at a selected randomization level to produce novel proteins. These proteins are then screened for members that are improved in their activity, specificity, affinity or antigenicity. This invention is exemplified with methods for randomizing antibody complementarity determining regions or framework regions and for enhancing the relatively low affinities that typify anti-carbohydrate antibodies.

Background and Prior Art

Protein engineering is a term used to define the manipulation of a protein, often by manipulating the DNA sequence to obtain proteins with a varied phenotype. Early techniques of mutating proteins involved manipulation of the amino acid sequence in a semi-synthetic method. In one example, the protein was cleaved into two fragments, an amino acid residue was removed from the new end of one of the fragments and a new amino acid was substituted. Other later techniques involved the manipulation of the protein at the DNA level by either random mutagenesis or at pre-selected positions by site-directed mutagenesis. The prior art techniques generally result in a single mutation per protein molecule.

Alternatively, mutations can be introduced by replacing a selected part of the wild-type gene sequence with random sequences or by gene amplification using PCR (Polymerase Chain Reaction) primers with randomized sequences. The limitation of the prior art methods is that they do not provide an efficient means of simultaneously and precisely targeting diversity to

multiple areas of genes. Using the prior art methods diversity can be introduced into only one or two regions of a gene.

In the present invention controlled simultaneous randomization of designated multiple areas of a DNA sequence is used to create libraries containing varying levels of amino acid substitution. In one example of the invention, the ligase chain reaction (LCR) process has been adapted to provide the adjoining of a number of oligonucleotides, together comprising a gene or portion thereof, wherein one or more targeted regions have been randomly and simultaneously mutated by spiking with the phosphoramidite derivatives of the four precursors of DNA: dA, dC, dG and T, at selected spiking levels. Using this process a library of randomized sequences can be been constructed.

The ligase chain reaction (LCR) has been previously used as a DNA diagnostic tool to covalently link synthetic oligonucleotides (Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) Science **241**, 1077-1080 and Barany, F. (1991) Proc. Natl. Acad. Sci, USA **88**, 189-193, both of which are incorporated herein by reference). The reaction exploits the capacity of a thermostable ligase to ligate two oligonucleotides which are hybridized to a targeted DNA sequence, provided that the adjoining ends are perfectly base-paired to the target.

The library of randomized sequences of the present invention, are expressed in a host system, which will preferably display the randomly generated proteins on the outer surface of a chosen host system. The phage display technology, which has recently been developed for the isolation of antibody fragments by expression of natural or semi-synthetic variable-gene libraries on

bacteriophage surfaces, represents one such host system (Barbas, C.F., Bain, J.D., Hoekstra, D.M. and Lerner, R.A. (1992) Proc. Nat'l. Acad. Sci. U.S.A. **89**, 4457-4461; Barbas, C.F., Kang, A.S., Lerner, R.A., and Benkovic, S.J. (1991) Proc. Nat'l. Acad. Sci. U.S.A. **88**, 7978-7982; Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J. Griffiths, A.D., and Winter, G. (1991) J. Mol. Biol. **222**, 581-597; and McCafferty, J., Griffiths, A.D., Winter, G., and Chiswell,D.A. (1990) Nature **348**, 552-554, all of which are incorporated herein by reference).

Summary of the Invention

The present invention relates to a method for precisely targeting randomization of a protein at a selected number of sites along the protein to produce novel proteins. Some examples of proteins that could be engineered using this invention are antibodies, hormones, receptor proteins, DNA binding proteins, RNA binding proteins and enzymes. These proteins are then screened for members that have improved properties. For example, these members may be improved in their activity, specificity, affinity or antigenicity.

Oligonucleotides for assembly of the gene, or a partial region thereof, to be mutated are synthesized using a DNA synthesizer, such that a portion of the nucleotide bases are randomized by controlled spiking with the phosphoramidite derivatives of the four precursors of DNA: dA, dC, dG and T. Some of the oligonucleotides so produced will be spiked at specific regions at specific levels of spiking, each parameter to be individually determined at the time of synthesis.

The oligonucleotides can then be joined using a ligase chain reaction (LCR) procedure. The LCR can be further modified by eliminating the requirement for

template DNA. In this example there may be a more stringent requirement for overlapping oligonucleotide ends that anneal efficiently. Alternatively, the randomized oligonucleotides can also be joined by a conventional T4 ligase reaction. Only the oligonucleotides of the correct size and base-pairing at adjoining ends are incorporated into the ligase products.

For efficient base pairing in the spiked regions, the sense strand can contain the randomized sequences while the antisense strand can contain inosine in these regions to act as a general base-pairing partner. The mutated full-length ligase products are then directly cloned into a display vector or amplified by PCR (polymerase chain reaction). The full-length mutated ligase product or the PCR-amplified product are then modified, if required, by restriction endonuclease digestion to ensure proper ligation with a host vector, for expression. The resulting library is then screened for desirable, mutated products.

Thus in one embodiment of the present invention there is provided a method for simultaneous controlled randomization of a protein at a number of sites along the protein to produce mutated proteins comprising the steps of:

synthesizing oligonucleotides of a gene, or a partial region thereof, encoding a protein by controlled randomized, by spiking with the phosphoramidite derivatives of the four precursors of DNA: dA, dC, dG and T, to produce oligonucleotides wherein a portion of said oligonucleotides are spiked at specific regions and at specific levels of spiking;

joining the oligonucleotides using a ligase reaction to produce a plurality of full-length mutated ligase products with two or more mutated regions;

ligating said products with a host vector for expression; and

screening for mutated proteins.

In a further embodiment of this invention the heavy chain portion of a wild-type gene encoding a single-chain antibody specific for Salmonella serogroup B O-polysaccharide was randomized.

In yet a further embodiment of the present invention, the single-chain antibody specific for Salmonella serogroup B O-polysaccharide was randomized in all of the heavy and light chain complementarity-determining regions (CDRs).

Further according to the present invention there is provided a method for randomizing the framework residues in a single chain antibody for the generation of a "humanized" framework library.

Brief Description of the Drawings

Figure 1: Predicted levels of amino acid substitutions at different spiking levels for the randomization of amino acid residues at a total of 19 positions.

Figure 2: Simultaneous randomization of three CDRs by LCR using spiked oligonucleotides. The bold line represents the template DNA for LCR and the double arrows indicate the DNA sequences encoding the five, eight and six amino acid stretches (in CDR1, CDR2 and CDR3, respectively) that were mutated by the spiking procedure. P1 and P2 represent the PCR primers used for product detection and amplification.

Figure 3: Agarose gel electrophoresis of PCR

and LCR products from oligonucleotides encoding randomized Se155-4 V_H CDRs. Lane 1, PCR product from LCR mixture at step 1; lane 2, final LCR product at position indicated by arrowhead; lane 3, Bethesda Research Laboratory 123-bp ladder. The lane 1 PCR product is slightly larger than the lane 2 LCR product because of the restriction site flanking sequences in the P1 and P2 primers.

Figure 4: Heavy chain sequences of single chain antibody clones, randomly picked before biopanning, showing mutations clustered in the three CDRs. The sequences subjected to 10% spiking are enclosed in boxes. The eight clones contained all possible base substitutions (i.e. A to T, G, C; G to T, A, C; T to A, C, G; C to G, T, A). Some deletions (indicated by D) and framework mutations, thought to be DNA synthesizer errors, were also observed. Both the sense and anti-sense strands are shown for the wild-type sequence but only the sense strand is given for the mutant sequences with the dots representing unmutated nucleotides.

Figure 5: Sensorgrams showing biphasic (A) and monophasic (B) binding of scFvs to immobilized BSA-O-polysaccharide. In the biphasic example, the wild-type was assayed at concentrations of 100 nM, 200 nM, 500 nM and 700 nM. In the monophasic example mutant B5-5 was assayed at concentrations of 20 nM, 40 nM, 80 nM, 150 nM and 200 nM.

Figure 6: Antigen-antibody complex showing the region where unfavourable contacts exist between the wild-type heavy chain CDR2 loop and the extended epitope. The antigen is a heptasaccharide in which the abequose has been removed from the second repeating unit.

Figure 7: Two-stage scFv library generation strategy showing the LCR assembly of eight oligonucleotides encoding the CDR randomized V_H library followed by the LCR assembly of ten oligonucleotides encoding the CDR randomized V_L library using the wild-type (WT) scFv gene as the template DNA in the first stage and mutant B5-1, isolated from the first stage library, as the template DNA in the second stage.

Figure 8: Agarose gel electrophoresis of T4, LCR and PCR products described in Example 3. Lane 1, BRL (Bethesda Research Laboratory) 100-bp ladder; lane 2, PCR product from final LCR product; lane 3, final LCR product; lane 4, PCR product from T4 product; lane 5, T4 product. Ten percent of the T4, LCR and PCR products were applied to the gel in each instance.

Detailed Description of the Preferred Embodiment

This invention relates to a method for the controlled randomization of a gene encoding a protein at a selected number of sites along the gene and at a selected randomization level to produce novel proteins. Some examples of proteins that could be engineered using this invention are antibodies, hormones, receptor proteins, DNA binding proteins, RNA binding proteins and enzymes. If structural information is available, regions of the protein that contact the ligand or substrate would, in most instances, be targeted. These novel proteins are then screened for members that are improved, for example, in their activity, specificity, affinity or antigenicity.

Oligonucleotides of a gene, or a partial region thereof, to be mutated, are synthesized using a DNA synthesizer, such that a portion of the nucleotide bases to be mutated is randomized by spiking with the phosphoramidite derivatives of the four precursors of

DNA: dA, dC, dG and T, at a controlled spiking level. The oligonucleotides are then joined using a ligase reaction. For example the oligonucleotides can be joined using the ligase chain reaction (LCR) procedure to produce full-length mutated ligase products with two or more mutated regions. The LCR reaction can be further modified by eliminating the requirement for template (wild type) DNA. In such instances, there may be a more stringent requirement for overlapping oligonucleotide ends that anneal efficiently. Alternatively, the randomized oligonucleotides can also be joined by a conventional T4 ligase reaction. Only the oligonucleotides of the correct size and base-pairing at adjoining ends are incorporated into the ligase products.

For efficient base pairing in the spiked regions, the sense strand can contain the randomized sequences while the antisense strand can contain a non-specific base-pairing partner, such as inosine, in these regions to act as a spacer. The mutated full-length ligase products are then directly cloned or amplified by PCR (polymerase chain reaction). The full-length mutated ligase product and the PCR-amplified product may require modification to ensure proper ligation with a host vector for expression. The resulting library is then screened for desirable, mutated products.

The equation

$$P=\frac{n!}{k!\ (n-k)!}\ m^k(1-m)^{n-k}$$

where

P = probability of a specified number of residue substitutions

m = mutation frequency of each amino acid at a

selected spiking level, ignoring triplet redundancy

k = number of amino acid substitutions obtained, and

n = number of randomized residues

was used to predict the probabilities of residue substitutions at different spiking levels.

Once the degree of randomization has been determined using the equation given above, sense strand oligonucleotides of variable length, for a full length gene, or a partial region thereof, are synthesized, such that some of the oligonucleotide are randomized by spiking at a level ranging from 2 - 90%. Spiking levels of 10%-70% and 20%-40% are also useful. However, in general the level of spiking will vary, depending on the individual situation and the desired frequency of amino acid substitution.

The oligonucleotides are synthesized with an automatic DNA/RNA synthesizer, with the phosphoramidite derivatives of the four precursors of DNA: dA, dC, dG and T, as phosphoramidite derivatives, using the following eight reagents: (i) 100% dA, (ii) 100% dC, (iii) 100% dG, (iv) 100% T, (v) x% dA + (100-x)% of an equimolar mixture of dA, dC, dG and T, (vi) x% dC + (100-x)% of an equimolar mixture of dA, dC, dG and T, (vii) x% dG + (100-x)% of an equimolar mixture of dA, dC, dG and T, (viii) x% T + (100-x)% of an equimolar mixture of dA, dC, dG and T; wherein 100-x equals the degree of randomization required. In general some of the oligonucleotides will contain spiked regions whereas other oligonucleotides will be unspiked. The degree of spiking and the position of spiking is under the control of the skilled person, practicing the present invention and will varying depending on the degree of amino acid substitution required.

The antisense strand oligonucleotides are also synthesizing with an automatic DNA/RNA synthesizer, such that the spiked regions of the sense oligonucleotides are paired with inosine in the antisense strands, using the five phosphoramidite derivatives: dA, dC, dG, T and dI using the following five reagents: (i) 100% dA, (ii) 100% dC, (iii) 100% dG, (iv) 100% T and (v) 100% dI. The synthetic oligonucleotides were prepared according to standard procedures, for example as published in Efcavitch, J.W., "Automated System for the Optimized Chemical Synthesis of Oligodeoxyribonucleotides" in Macromolecular Sequencing and Synthesis. Selected Methods and Applications. pages 221-234, Alan R. Liss, Inc. (1988), incorporated herein by reference.

In one embodiment of the present invention, the oligonucleotides were joined using the ligase chain reaction (LCR) procedure as described by Landegren, U., Kaiser, R., Sanders, J. and Hood, L. (1988) Science **241**, 1077-1080 and Barany, F. (1991) Proc. Natl. Acad. Sci, USA **88**, 189-193, both of which are incorporated herein by reference to, produce mutated full-length LCR products. The LCR procedure was modified to allow assembly of up to ten oligonucleotides by performing the ligation reaction, using a thermostable DNA ligase and a DNA template, in a thermocycler in multiple steps and cycles which may vary depending on the reactants.

If desired, the LCR reaction can be further modified by eliminating the requirement for template (wild type) DNA. In such instances, there may be a more stringent requirement for overlapping oligonucleotide ends that anneal efficiently.

In a further embodiment of the present invention the standard T4 ligation reaction (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular

Cloning: A Laboratory Manual (2nd ed), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., U.S.A.) was also used.

Regardless of the method used for the ligation of the oligonucleotides, the ligation reaction will produce full-length mutated ligase products with two or more mutated regions.

If mutated full-length ligase product yield is insufficient, the product can be amplified by PCR (polymerase chain reaction), employing standard PCR reaction conditions, for example as discussed in Mullis, K.B., and Faloona, F.A. (1987) Methods Enzymol. **155**, 335-350, incorporated herein by reference.

The full-length mutated ligase product and/or the PCR-amplified product can then be modified, if required, by restriction endonuclease digestion to ensure proper ligation with a host vector for expression. The restriction endonuclease used will depend on the host vector chosen for expression and the specific cloning sites in the chosen vector. There are a number of suitable vectors that can be used according to the present invention, either phage-based or plasmid-based vectors. The choice of host vector, and the modification of the ligase product or the PCR-amplified product required to ensure proper ligation in a host vector are well within the knowledge of a reasonably skilled person in this area of technology.

Cloned segments, which bind to a particular ligand or substrate, can be selected from large phage display libraries by a procedure referred to as "biological panning" or "panning", (Deng, S., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994), J. Biol. Chem.,

269, 9533-9528). In this procedure, the chosen ligand or substrate is coated on the surface of a solid support tube or plastic dish. An aliquot of the phage library is added to the tube. Those phage that can bind to the ligand or substrate become indirectly immobilized to the wall of the tube. Non-binding phage are subsequently washed from the tube or dish and the phage expressing a binding segment are eluted by a short treatment at high or low pH. The recovered phage can then be amplified using standard techniques. Phage expressing products with altered affinities, specificities or binding kinetics, relative to the wild-type protein are further screened by colony blot, ELISA or biosensor procedures that measure binding of soluble binding protein or phage-displayed binding protein to a chosen ligand or substrate. Novel proteins that are improved in their activity, specificity, affinity or antigenicity are thus selected.

Two of the examples given herein are directed to the <u>in vitro</u> affinity maturation of an antibody. The six CDRs are randomized such that amino acid substitutions are introduced at a level that approximates that resulting from somatic hypermutation during the affinity maturation stage of the immune response.

The present invention could also be used to humanize rodent antibodies so they can be used in therapy. In one example given herein, selected residues in human frameworks are randomized, generating libraries for the selection of fully active humanized antibodies. Generally the use of a mouse antibody in therapy results in a human anti-mouse antibody response, which neutralizes the mouse antibody. Simple grafting of mouse CDRs on human frameworks generally gives a humanized form that is much less active than the mouse

antibody because of altered CDR conformations with the new framework scaffolding. As an alternative to rational redesign of the CDR-grafted antibodies to restore full activity, the present randomized library approach, in which the limited number of residues which differ in the mouse and human frameworks are randomized, in conjunction with phage display, provides a workable approach. As the residues to be randomized are scattered throughout the molecule, the ligase chain reaction described in the present invention would provide the means of generating such libraries. Therefore not only can the CDRs be randomized but also the frameworks. In another approach to humanize rodent antibodies, only the exposed residues of the murine antibodies would be randomized, resulting in resurfaced murine antibodies.

Single domain antibodies, for example the V_H domain are of current interest because of their small size. However, these antibodies have relatively low activity and poor solubility because of the exposed hydrophobic surface that is normally associated with the V_L domain. The simultaneous randomization method of the present invention can be used to improve the properties of the single domain antibodies. By randomizing the residues that contribute to the hydrophobic surface, displaying the libraries on phage and panning against antigen, molecules with better biological properties, such as solubility, can be selected. The modified LCR technology of the present invention provides an efficient means of generating these libraries.

In practicing the present invention, there will be many mutants that will no longer bind the original antigen. These mutants may, however, have specificities for other antigens and panning these libraries against various antigens may provide an easy

way of generating new antibodies including some that the immune system cannot make, for example against toxic substances.

While this invention is described in detail with particular reference to preferred embodiments thereof, said embodiments are offered to illustrate but not limit the invention.

Examples

Example 1. One-stage in Vitro Affinity Maturation of a Single-chain Antibody.

The antibody, Se155-4, is specific for the Salmonella serogroup B lipopolysaccharide antigen, which is built from the four sugar repeating unit:

{→2)[αD-Abe(1→3)]αD-Man(1→4)αL-Rha(1→3)αD-Gal(1→}.

The trisaccharide epitope Gal-[Abe]-Man is bound in a "pocket"-like site via a network of hydrogen bonds and van der Waals contacts. Abequose, the 3,6-dideoxy immunodominant sugar, is totally buried while mannose and galactose residues lie on the protein surface and are partially exposed to water.

In this example k, from the equation discussed above is from 0 to 19 and n equals 19. The probabilities of residue substitutions at different spiking levels are shown in Figure 1.

Figure 2 is a pictorial representation of the simultaneous randomization procedure showing the eight oligonucleotides used in Example 1.

For the sense strand, 56 to 85 mer oligonucleotides were synthesized using an ABI 384 DNA/RNA synthesizer with 15 to 24 of the central bases, encoding CDRs, spiked at a level of 10% with an

equimolar mixture of the phosphoramidite derivatives of the four nucleosides. The antisense oligonucleotides were synthesized so that the spiked regions were paired with inosine. The LCR mixture contained 4pmol of each of the eight oligonucleotide building blocks for the variable heavy chain (V_H) (12) in a total volume of 8 µl, 5 µl 10x reaction buffer (200mM Tris-HCl, pH7.6, 100mM KCl, 100mM $MgCl_2$, 1% Triton X-100, 1mM ATP, 10mM DTT), 4 µl Pfu DNA ligase (Stratagene), 5 µl (20 ng) template DNA to facilitate ordered ligation of the multiple oligonucleotides and 30 µl distilled water. LCR was performed in a thermocycler (Perkin Elmer Cetus GeneAmp PCR System 9600) by running step I (92°C for 3 min, 60°C for 3 min) for 1 cycle, step II (92°C for 45 sec, 60°C for 30 sec) for 30 cycles and step III (92°C for 4 min, 60°C for 3 min) for 1 cycle. The 300-bp LCR product was purified by electrophoresis and removed from the agarose gel using standard procedures (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (2nd ed), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., U.S.A.). A range of product sizes was observed throughout the LCR procedure but full-length product was present as early as step I and was a significant component of the mixture at step III (Figure 3). After phosphorylation (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (2nd ed), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., U.S.A.), half of V_H product was directly cloned into Nhe I-Bgl II fragment of the wild-type phagemid vector, pSK4, (Deng, S., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994), J. Biol. Chem., **269**, 9533-9528) and half was PCR-amplified using upstream primer 5'-AGCTGCAAAGCTAGCGGTTACACCTTCACC-3' and downstream primer 5'-CGCCACCAGATCTGGAGGACACGGTCAGGCTCGCGCCTTGG-3', the latter containing a Sty I site which is not present in

the wild type V_H. PCR was carried out using Taq DNA polymerase in a Geneamp 9600 system (Perkin Elmer Cetus) at 94°C for 3 min (1 cycle), 94°C for 35 sec, 50°C for 45 sec and 72°C for 30 sec (15 cycles) and 72°C for 5 min for the final extension. The PCR-amplified product, following NheI-BglII digestion was cloned into Nhe I-Bgl II fragment of the phagemid vector pSK4 containing the light chain segment of the scFv gene. Phagemid preparations containing the direct LCR product and the PCR-amplified LCR product were pooled; this was done in an effort to give sufficient yield while maintaining any library diversity that might have been lost in the PCR step. The resulting mutated phagemid library was electroporated into XL1-Blue cells, resulting in 2×10^7 transformants.

Randomly picked clones from the unpanned library were sequenced using a ds DNA cycle sequencing kit (Life Technologies Inc.) with 5'-TGCGAGCGT-TAAAATGAGCTGC-3' and 5'-CGATTGGCCTTGATATTCACAAACG-3' as primers. The randomly picked clones showed a clustering of mutations in the heavy chain CDRs with no obvious bias for particular base substitutions, at a frequency that was in agreement with the predicted levels (Table 1 and Figure 4). Some frame shift mutations were also observed.

TABLE 1

Predicted and observed mutation frequencies in heavy chain CDR-randomized (10%) library.						
	k					
	0	1	2	3	4	5
Predicted cfu	4.8×10^5	1.5×10^6	2.4×10^6	2.3×10^6	1.5×10^6	7.6×10^5
Predicted P	5%	16%	25%	24%	16%	8%
Observed in unpanned clones	5%	20%	30%	25%	15%	5%
Observed in panned clones	10%	70%	20%	0	0	0

Phage were prepared from overnight cultures (Deng, S., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994), J. Biol. Chem., **269**, 9533-9528). Phage preparations were panned against Salmonella serogroup B LPS which was coated on microtitre plates at a concentration of 10 μg/ml. Clones randomly picked after three, four and five rounds of panning were screened for the diagnostic StyI site with the positive clones being further screened by an ELISA procedure in which antibody binding to plates coated with serogroup B LPS was detected with an anti-mouse lambda chain/alkaline phosphatase conjugate (Deng, S., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994), J. Biol. Chem., **269**, 9533-9528).

Clones displaying the highest activities were sequenced as described above. Nine sequences emerged from the fifty clones that were sequenced (Table 2).

TABLE 2

Amino acid differences of heavy chain mutants selected from a CDR-randomized library and generated by site-directed mutagenesis.

scFv type	Clone	Position								
		34*	43	46	55**	56**	57**	77	104***	109
wild-type	SK4	M	Q	E	N	S	A	I	Y	G
V_H library mutants	B3-13	I	-	-	-	-	-	T	-	-
	B3-19	I	-	-	-	-	-	-	-	-
	B3-20	-	-	V	Y	G	-	T	-	-
	B4-3	-	-	-	-	N	-	-	-	S
	B5-1	I	-	-	-	-	-	-	-	S
	B5-5	-	-	-	-	G	-	-	-	S
	B5-6	-	-	-	-	G	-	T	-	-
	B5-8	-	-	-	-	-	S	-	F	S
	B5-12	-	ΔQ	-	-	G	-	-	-	S

*CDR1, **CDR2, ***CDR3

Δ = deletion

For expression studies, mutants were digested with BglII for insertion of a self-complementary terminator sequence (Deng, S., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994) J. Biol. Chem., **269**, 9533-9528), between the scFv and gIII regions. Plasmids encoding wild-type and mutant scFvs were transformed into E. coli TG-1 for production of secreted and functional scFvs. Cultures were grown and scFvs were isolated from periplasmic extracts by affinity chromatography (Deng, S., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994) J. Biol. Chem., **269**, 9533-9528).

The kinetics of scFv binding to BSA-O-polysaccharide conjugates were determined by surface plasmon resonance using a BIAcore™ biosensor system (Pharmacia Biosensor). Immobilizations were

carried out in 10 mM Na acetate, pH 4.5, using the amine coupling kit supplied by the manufacturer and at concentrations and contact times that gave approximately 200 RU of immobilized conjugate. All measurements were carried out at 25°C in 10 mM HEPES, pH 7.4, 150 mM NaCl, 3.3 mM EDTA at a flow rate of 3µl/min. Surfaces were regenerated with 10 mM HCl. Association and dissociation rate constants were calculated using BIAlogue™ software (Pharmacia Biosensor). The affinity and kinetic results are shown in Table 3.

TABLE 3

Affinities and kinetics of BSA-O-chain binding by wild-type and mutant scFvs.

ScFv	k_{on} $M^{-1}s^{-1}$	k_{off} s^{-1}	K_d nM	K_d^*/K_d^{**}
WT	4.8×10^3	2.4×10^{-3}	500	1
B3-13	9.5×10^3	2.7×10^{-3}	283	1.8
B3-19	1.8×10^5	2.7×10^{-3}	13	50
B3-20	1.6×10^4	3.0×10^{-3}	191	2.6
B4-3	2.2×10^5	1.9×10^{-3}	8.5	58
B5-1	2.2×10^5	2.4×10^{-3}	11	42
B5-5	1.9×10^5	2.4×10^{-3}	12	42
B5-6	4.8×10^4	3.1×10^{-3}	64	7.8
B5-8	1.5×10^5	5.5×10^{-4}	3.6	139
B5-12	1.8×10^5	2.5×10^{-3}	14	36

*wild-type scFv; **mutant scFv

The K_d values obtained for most of the mutants were in the range of 30-40 fold that of the wild-type with the improved affinities being entirely attributable to faster on-rates (Table 3). In only one instance did a slower off-rate also contribute to higher affinity, giving a mutant with approximately 70-fold improved K_d relative to that of the wild-type. While the association phase of wild-type binding was distinctly biphasic, that of the mutants was classical (Figure 5).

The effect on binding of single residue substitutions at positions 34 and 109 was particularly striking. Neither of these are antigen-contacting residues. Position 34 is adjacent to His35H, a key HCDR1 residue that participates in a hydrogen bond network involving a water molecule at the base of the binding pocket. It is possible that the Met35HIle mutation may fine tune these interactions. The Gly109Ser mutation appears to allow for the formation of hydrogen bonds from Ser OH to the carbonyl oxygen of heavy chain residue 4 and the NH of heavy chain residue 6. The latter two residues form a beta bulge and the additional hydrogen bonds may stabilize the structure in the region where the linker approaches the VH domain.

Several of the mutants with improved binding properties had heavy chain CDR2 mutations which may remove some of the clashes between this loop and the extended epitope (Figure 6).

The Ile77HThr mutation was previously observed to have been selected on the basis of improved production of functional scFv in Escherichia coli (Deng, S., MacKenzie, C. R., Sadowska, J., Michniewicz, J., Young, N. M., Bundle, D. R., and Narang, S. A. (1994) J. Biol. Chem., **269**, 9533-9528).

Example 2. Two-stage in Vitro Affinity Maturation of a Single-chain Antibody.

Using similar reasoning and methods to those outlined in Example 1, in vitro affinity maturation of Se155-4 was performed in two stages. In the first stage, a mutant with improved binding was isolated from a heavy chain CDR-randomized library as described in Example 1. In the second stage, a light chain CDR-randomized LCR product is inserted into a phagemid

vector containing the heavy chain portion of the mutant single-chain antibody selected from stage 1.

Figure 7 is a pictorial representation of the two-stage affinity maturation process.

An LCR-generated V_L library containing a diagnostic BssHII site was constructed in a manner identical to that described in Example 1 except that 5 pmol of each of ten oligonucleotide building blocks for the V_L domain were used. Following digestion with XhoI and SacI, the V_L product was cloned into the XhoI-SacI fragment of the phagemid vector encoding mutant 5B-1, a V_H mutant (Met34Ile, Gly109Ser) isolated from the first stage library. The ligation products were electroporated into E. coli XL1-Blue, resulting in $2x10^7$ transformants. Phage were prepared and panned as in Example 1. Following panning, clones displaying higher activity than 5B-1 in the ELISA screening procedure described in Example 1 were sequenced.

Two sequences, one with a single CDR2 mutation and another with two CDR3 mutations, emerged from the 8 clones that were sequenced (Table 4). Relative to their 5B-1 predecessor, mutant 2-5 displayed a slightly faster on rate while mutant 1-4 had a somewhat slower off-rate (Table 5).

TABLE 4

Amino acid differences of mutants selected from a light CDR-randomized library.						
scFv type	Clone	Position				
		54L**	94L***	96L***	34H*	109H
wild-type	SK4	N	S	N	M	G
B5-1		-	-	-	I	S
V_L library mutants	1-4	G	-	-	I	S
	2-5	-	Y	D	I	S

*CDR1, **CDR2, ***CDR3

TABLE 5

Affinities and kinetics of BSA-O-chain binding by wild-type and mutant scFvs.					
scFv type	clone	k_{on} M^{-1} s^{-1}	k_{off} s^{-1}	K_d nM	K_d^*/K_d^{**}
wild-type	SK4	4.8×10^3	2.4×10^{-3}	500	1
B5-1		2.2×10^5	2.4×10^{-3}	11	42
V_L	1-4	1.3×10^5	9.4×10^{-4}	7	71
mutants	2-5	1.6×10^5	2.0×10^{-3}	12	42

Example 3. Randomization of selected framework residues in a single chain antibody for generation of a humanization library

A portion of Sel55-4 V_H was synthesized by LCR such that the murine framework 1, 2 and 3 regions were replaced with partially randomized frameworks from the human antibody NEW. The residues selected for

randomization were those which frequently differ in murine and human V_H frameworks. Six oligonucleotides (Table 6) encoding the targeted V_H frameworks were synthesized such that the positions shown in bold in Table 6 contained inosine or were randomized by spiking at a level of 10% with an equimolar mixture of the phosphoramidite derivatives of the four DNA precursors.

TABLE 6

Oligonucleotides used in construction of randomized framework library. Ten per cent randomization or inosine substitution was introduced at the positions shown in bold.

Oligonucleotide	Sequence
1	5′-CAGCTGGAACAGTCCGGC**CCTGG**CCTGGTT-CGCCCG**TCC**CAGACCCTG**TCCCTC**ACCTGCACC-GTTTCC-3′
2	5′-GGCACCTCCTTCAACAACTACTGGATGCAC-TGGGTT**CGC**CAG**CCTC**CGGGC**CGC**GGCCTGGAA-TGGATCGGCGCTATCTACCCG-3′
3	5′-GGCAACTCCTCCACCTTCTACAACCACAAA-TTC**GGCCGCGTT**ACCATGCTG**GTTAA**CACCTCC-**AAAAA**CCAGTTCTCCCTCCGCCTGTCCTCCGTG-ACC-3′
4	5′-GGTCACGGAGGACAGGIIGIIGIIGIICTG GIITTTGGAGGTGIIAIICAGCATGGTAIIGII-GIIGAATTTGTGGTTGTA-3′
5	5′-GAAGGTGGAGGAGTTGCCCGGGTAGATAGC-GCCGATCCATTCCAGGCCGIIGCCCGGAIICTG-GIIAACCCAGTGCATCCA-3′
6	5′-GTAGTTGTTGAAGGAGGTGCCGGAAACGGT GCAGGTGIIGIICAGGGTCTGGIICGGGCGAAC-CAGGIIAIIGCCGGACTGTTCCAGCTG-3′

Standard T4 ligation (Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (2nd ed), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., U.S.A.) and synthetic LCR, without template DNA, were investigated

as means of correctly assembling the oligonucleotides to form the partial V_H sequence. In each instance, 50 pmol of each of the six oligonucleotides (Table 6) were added to the reaction mixture. Apart from the absence of template DNA and the higher oligonucleotide concentrations, LCR conditions were as described in Example 1.

The T4 reaction gave one major band of the size expected for the correctly assembled partial V_H sequence (Fig. 8). However, PCR amplification of the T4 product showed two bands indicating non-specific ligation by T4 DNA ligase. The LCR reaction gave a higher yield of product that was shown by PCR to consist of only the correctly assembled sequence indicating highly specific ligation. In both instances, PCR extended the product to introduce suitable restriction sites at the 5′ and 3′ ends.

The presence of correctly ligated T4 product was not observed with the oligonucleotides described in Examples 1 and 2. The presence of only six oligonucleotides is thought to be the reason for its formation in this instance.

Replacing the wild-type murine sequence with the LCR product thus obtained should generate a library from which fully active humanized versions of Se155-4 V_H could be selected by the phage display techniques described in Example 1.

Various modifications may be made to the preferred embodiments without departing from the spirit and scope of the invention as defined in the appended claims.

THE EMBODIMENTS OF THE INVENTION IN WHICH AN EXCLUSIVE PROPERTY OR PRIVILEGE IS CLAIMED ARE DEFINED AS FOLLOWS:

1. A method for simultaneous controlled randomization of a protein at a number of sites along the protein to produce mutated proteins comprising the steps of:

synthesizing oligonucleotides of a gene, or a partial region thereof, encoding a protein by controlled randomization, by spiking with the phosphoramidite derivatives of the four precursors of DNA: dA, dC, dG and T, to produce oligonucleotides wherein a portion of said oligonucleotides are spiked at specific regions and at specific levels of spiking;

joining the oligonucleotides using a ligase reaction to produce a plurality of full-length mutated ligase products with two or more mutated regions;

ligating said products with a host vector for expression; and

screening for mutated proteins.

2. A method according to Claim 1, wherein the full-length mutated ligase product is amplified by polymerase chain reaction (PCR) to produce a PCR-amplified product, and wherein said PCR-amplified product is modified to ensure proper ligation with a host vector for expression.

3. A method according to Claim 2, wherein the sense strand of the gene, or partial region thereof, is randomized by spiking to produce a mutated region.

4. A method according to Claim 3, wherein the anti-sense strand of the gene, or partial region thereof, contains a non-specific base pairing partner in the mutated region.

5. A method according to Claim 4, wherein the non-

specific base pairing partner is inosine.

6. A method according to Claim 1, wherein the gene or partial region thereof is randomized by spiking at a level of from 2 to 90 percent.

7. A method according to Claim 6, wherein the protein is selected from the group consisting of antibodies, hormones, receptor proteins, DNA-binding proteins, RNA-binding proteins and enzymes.

8. A method according to Claim 7, wherein the full-length mutated ligase product or amplified PCR product is modified by restriction endonuclease digestion to ensure proper ligation with a host vector.

9. A method according to Claim 8, wherein the host vector is selected from the group consisting of a phage-based vector or a plasmid-based vector.

10. A method according to Claim 9, wherein the mutated proteins are screened by biopanning and further selected by a method selected from the group consisting of colony blot, hybridization, ELISA, and biosensor procedures that measure binding of soluble binding protein or phage-displayed binding protein to a chosen ligand.

11. A method according to Claim 1, wherein the ligase reaction is selected from the group consisting of a ligase chain reaction, a modified ligase chain reaction, wherein said reaction is performed in the absence of template DNA, and T4 ligase reaction.

12. A method for simultaneous controlled randomization of a protein at a number of sites along the protein to produce mutated proteins comprising the steps of:

synthesizing oligonucleotides of a gene, or a

partial region thereof, encoding a protein selected from the group consisting of antibodies, hormones, receptor proteins, DNA-binding proteins, RNA-binding proteins and enzymes by controlled randomized, by spiking at a level of from 2 to 90 percent with the phosphoramidite derivatives of the four precursors of DNA: dA, dC, dG and T, to produce oligonucleotides wherein a portion of said oligonucleotides are spiked at specific regions and at specific levels of spiking;

joining the oligonucleotides using a ligase chain reaction (LCR) to produce a plurality of full-length mutated ligase products with two or more mutated regions;

modifying the full-length mutated ligase product, by restriction endonuclease digestion, to ensure proper ligation with a host vector for expression; and

screening for mutated proteins improved in their activity, specificity, affinity or antigenicity.

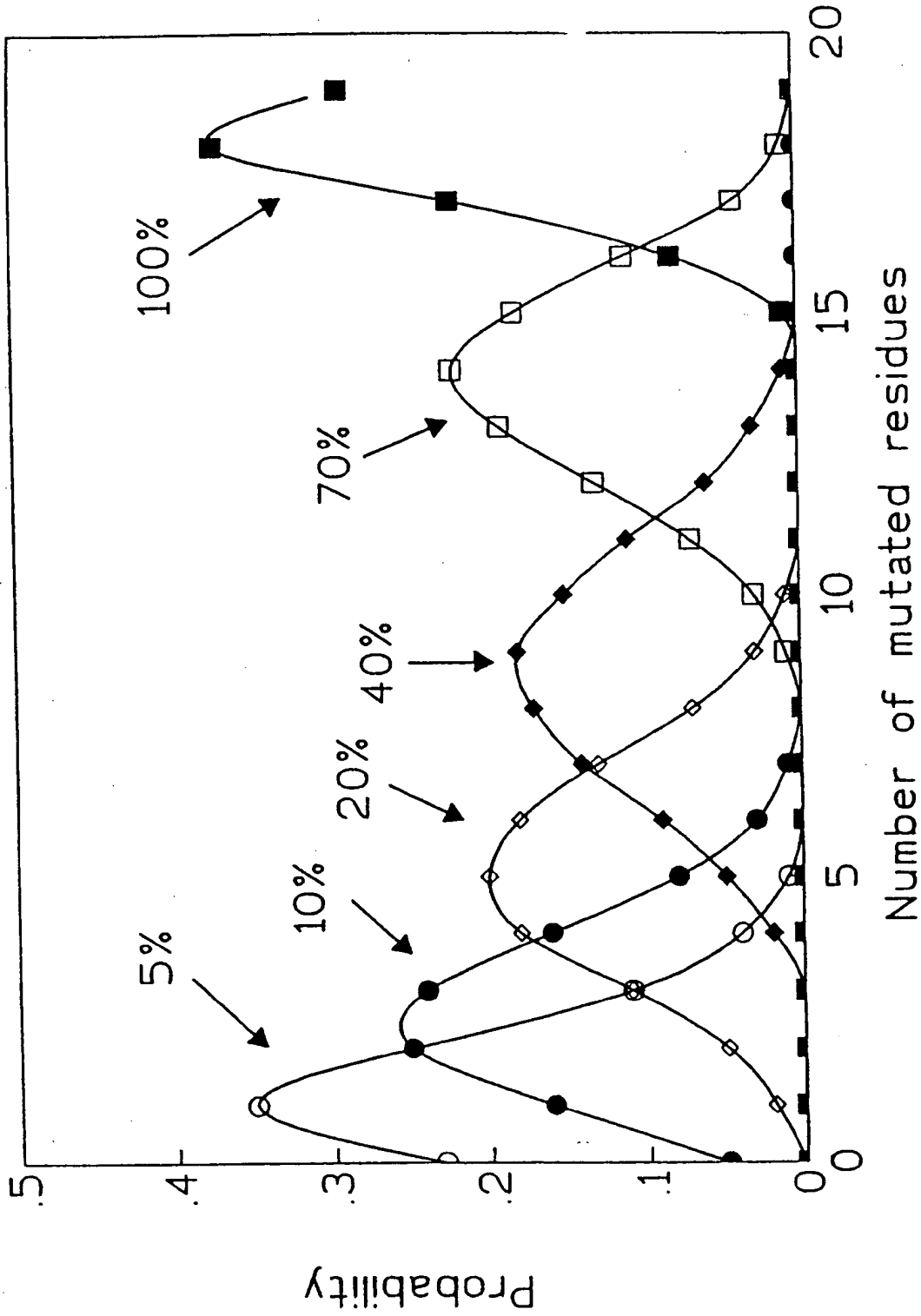
Probability
Number of mutated residues
0
5
10
15
20
.1
.2
.3
.4
.5
5%
10%
20%
40%
70%
100%

FIGURE 1

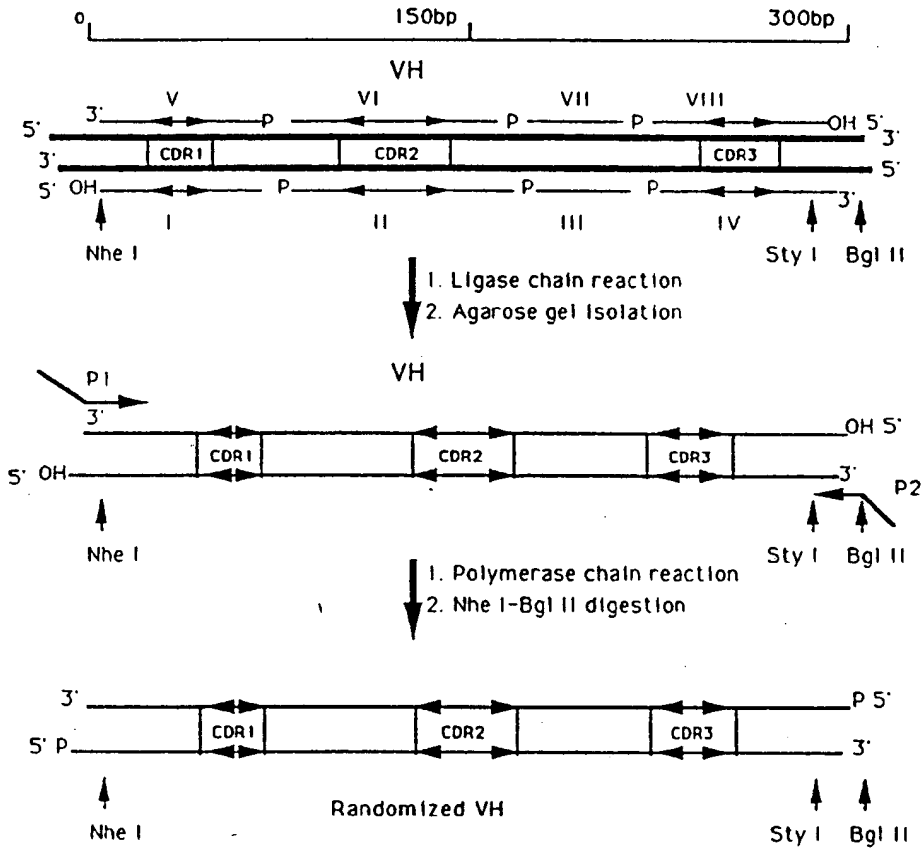
0
150bp
300bp
VH
V
VI
VII
VIII
3'
5'
P
OH 5'
3'
CDR1
CDR2
CDR3
5' OH
I
II
III
IV
Nhe I
Sty I
Bgl II
1. Ligase chain reaction
2. Agarose gel isolation
VH
P1
OH 5'
5' OH
P2
1. Polymerase chain reaction
2. Nhe I-Bgl II digestion
P 5'
5' P
Randomized VH

FIGURE 2

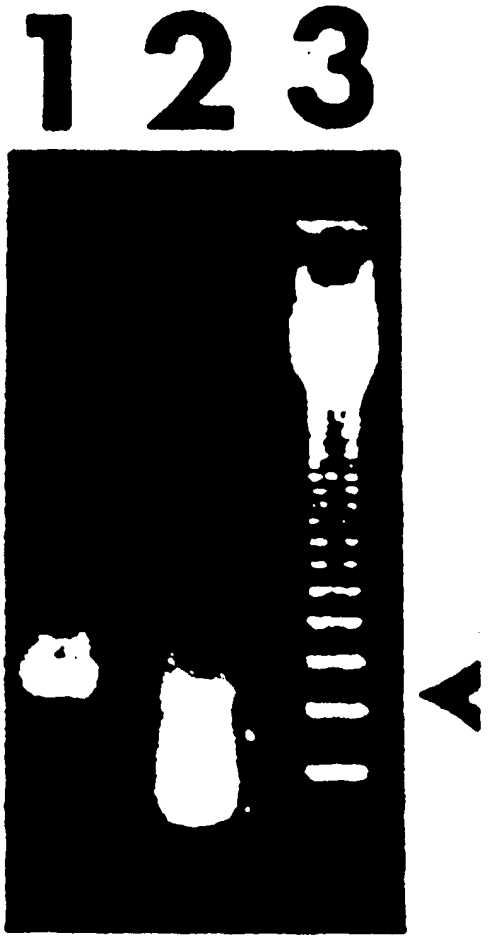

FIGURE 3

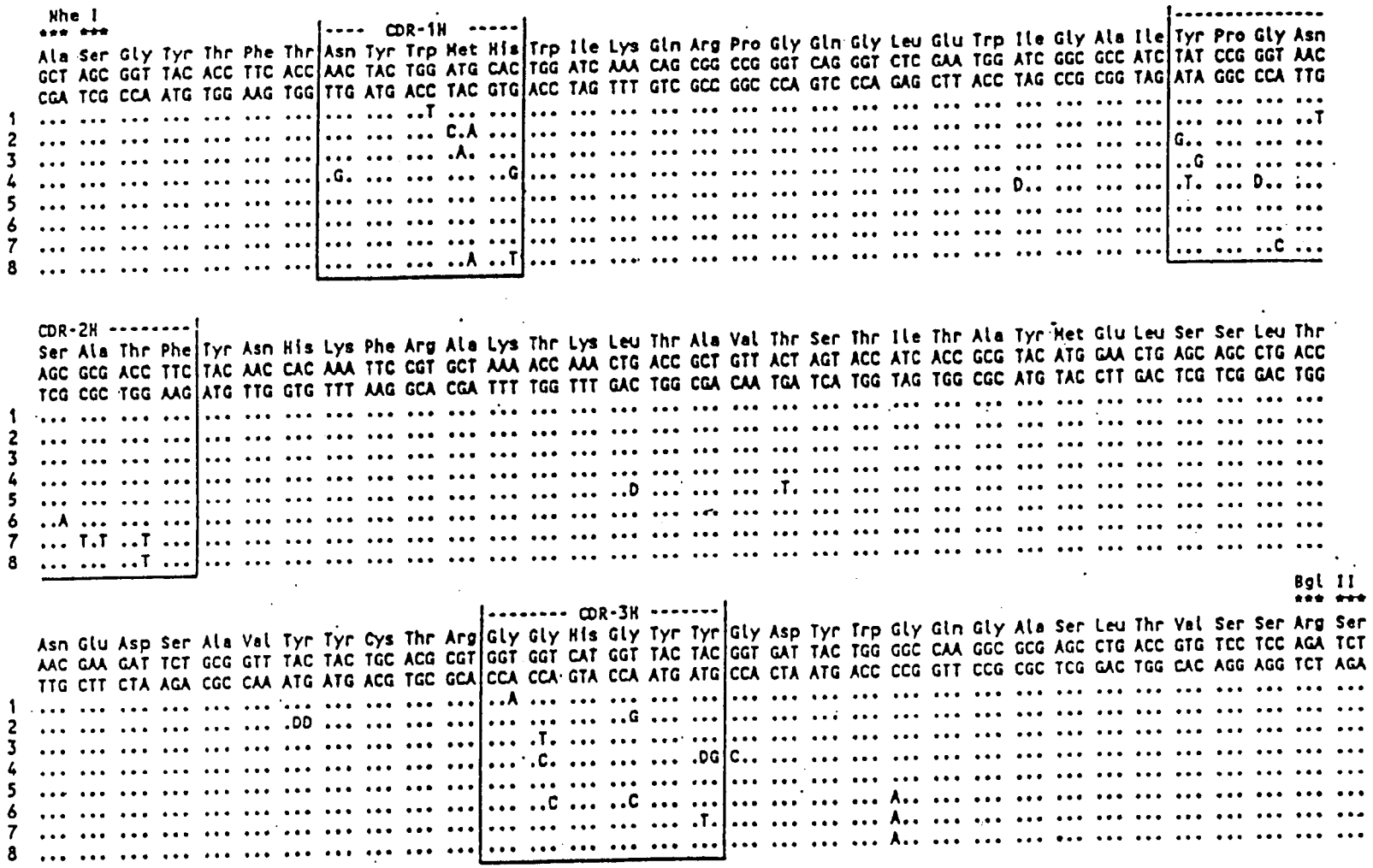

FIGURE 4

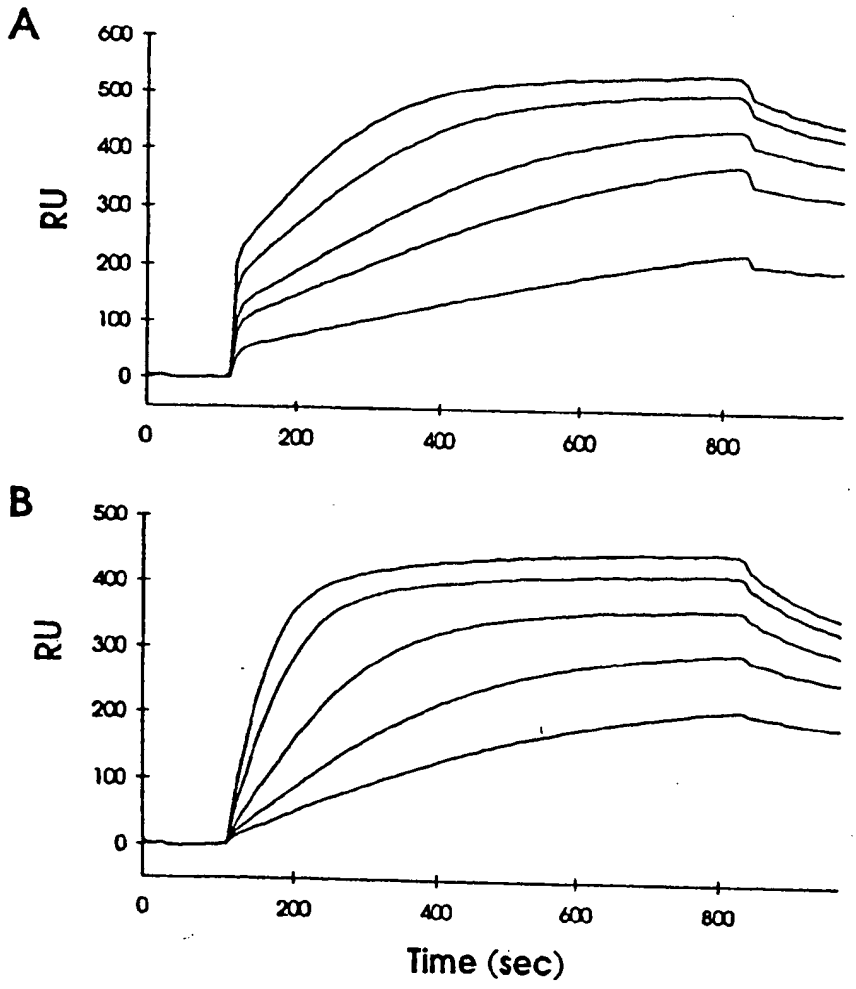
A
600
500
400
300
200
100
0
RU
0
200
400
600
800
B
500
400
300
200
100
0
RU
0
200
400
600
800
Time (sec)

FIGURE 5

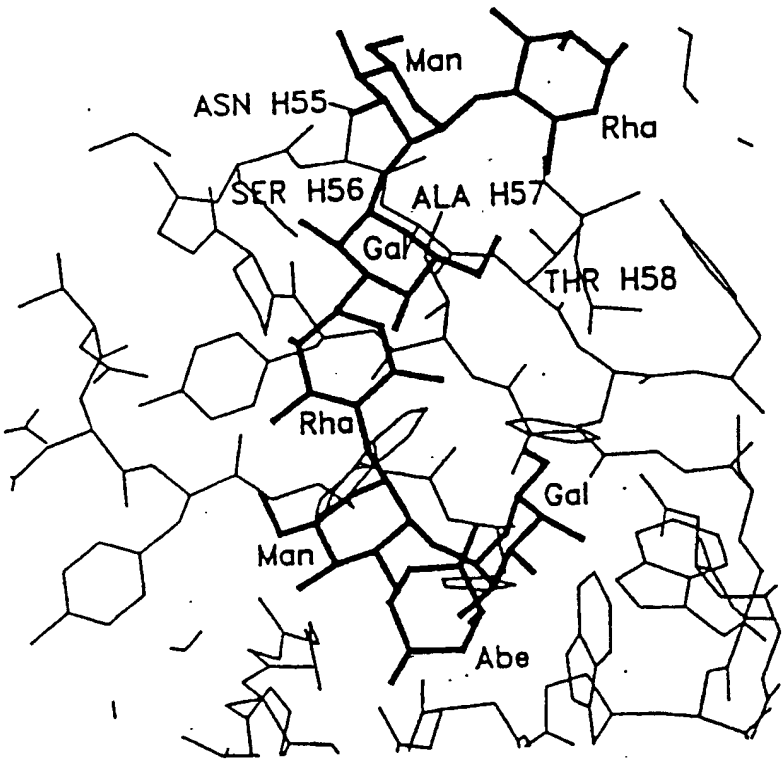

FIGURE 6

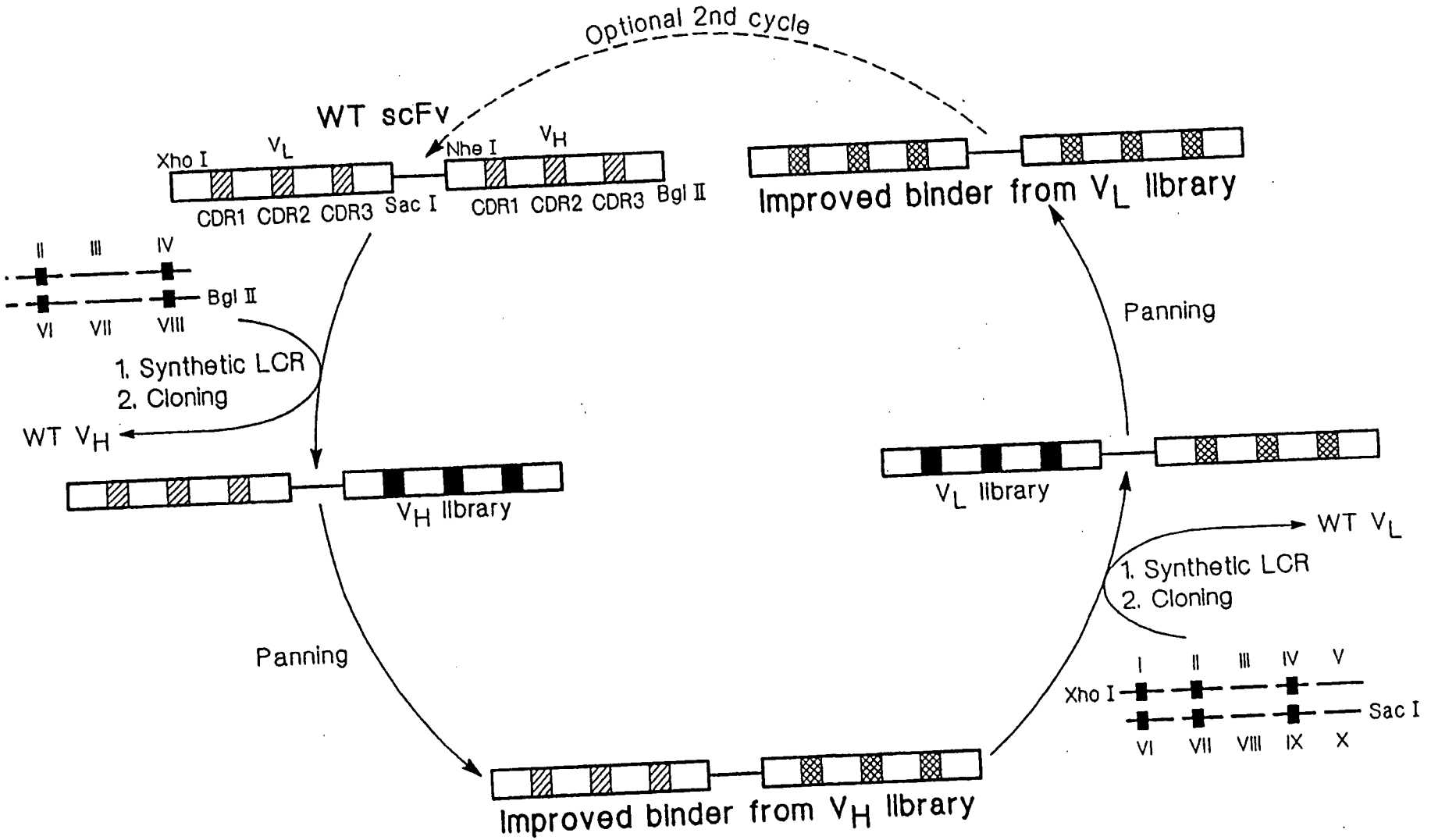

FIGURE 7

FIGURE 8